

MENU

SEARCH

INDEX

DETAIL

JAPANESE

BACK

NEXT

2 / 18

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-207712
 (43)Date of publication of application : 26.07.2002

(51)Int.Cl. G06F 15/177
 G06F 9/46
 G06F 13/00

(21)Application number : 2001-312183 (71)Applicant : HEWLETT PACKARD CO <HP>
 (22)Date of filing : 10.10.2001 (72)Inventor : ROMERO FRANCISCO J
 THOMAS E TURICCHI JR

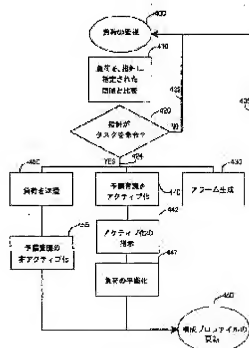
(30)Priority
 Priority number : 2000 709705 Priority date : 09.11.2000 Priority country : US

(54) DEVICE AND METHOD FOR ACTIVATING RESERVE RESOURCE AUTOMATICALLY

(57) Abstract:

PROBLEM TO BE SOLVED: To dynamically, i.e., automatically provide a resource capacity required to cope with variation of a user demand, and automatically activate a reserve resource when an additional resource capacity is required in order to satisfy the demand.

SOLUTION: This method for activating the reserve resource automatically contains a step 400 for monitoring load to some active resources, a step 410 for comparing the load with a threshold value assigned in a resource using plan, and a step 440 for activating the reserve resource automatically when a command is issued according to the resource using plan.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(51) Int.Cl. ⁷	識別記号	F I	テーマコード [*] (参考)
G 0 6 F 15/177	6 7 4	G 0 6 F 15/177	6 7 4 A 5 B 0 4 5
9/46	3 6 0	9/46	3 6 0 C 5 B 0 8 9
13/00	3 5 7	13/00	3 5 7 Z 5 B 0 9 8

審査請求 未請求 請求項の数1 O L (全 10 頁)

(21) 出願番号 特願2001-312183(P2001-312183)
(22) 出願日 平成13年10月10日 (2001.10.10)
(31) 優先権主張番号 0 9 / 7 0 9 7 0 5
(32) 優先日 平成12年11月9日 (2000.11.9)
(33) 優先権主張国 米国 (U S)

(71) 出願人 398038580
ヒューレット・パッカード・カンパニー
HEWLETT-PACKARD COM
PANY
アメリカ合衆国カリフォルニア州パロアル
ト ハノーバー・ストリート 3000
(72) 発明者
フランシスコ・ジェイ・ロメロ
アメリカ合衆国75025テキサス州プラノ、
アヴァロン・ドライブ 7609
(74) 代理人 100081721
弁理士 岡田 次生 (外2名)

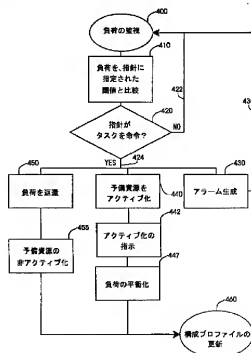
最終頁に続く

(54) 【発明の名称】 自動的に予備資源をアクティブ化する装置および方法

(57) 【要約】

【課題】 本発明の主な目的は、動的にすなわち自動的に、ユーザ需要の変動に応じるために必要な資源容量を提供することである。本発明の別の目的は、需要を満たすために、追加の資源容量が必要な場合に、予備資源を自動的にアクティブ化することである。

【解決手段】 予備資源を自動的にアクティブ化する方法であって、いくつかのアクティブ資源に対する負荷を監視するステップ (400) と、前記負荷を資源使用方針に指定されている閾値と比較するステップ (410) と、前記資源使用方針によって命令される場合に、前記予備資源を自動的にアクティブ化するステップ (440) と、を含む。



1

【特許請求の範囲】

【請求項1】 予備資源を自動的にアクティブ化する方法であって、

いくつかのアクティブ資源に対する負荷を監視するステップと、

前記負荷を資源使用方針に指定されている閾値と比較するステップと、

前記資源使用方針によって命令される場合に、前記予備資源を自動的にアクティブ化するステップと、を含む、方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、いくつかのアクティブ資源の負荷を監視し、いくつかのアクティブ資源の負荷が資源使用方針において指定された閾値に見合う場合に、予備資源を自動的にアクティブ化する装置および方法に関する。

【0002】

【従来の技術】 計算およびネットワーク資源は、需要に応じて広く変動する。たとえば、計算システムまたはそのコンポーネント（たとえば、CPUまたはメモリ）は、早朝時や週末での負荷がわずかにあるか、ない状態から、週日の夕方近くの使用や、または多くの計算システムが大容量の計算および他のデータ処理を行うようプログラムされている夕刻での非常に高い負荷までの変動を受ける可能性がある。同様に、インターネットまたは他のネットワーク資源（たとえば、サーバやホストコンピュータ）に対するユーザの需要は、ユーザ数または時刻等、様々な要因に依存する。さらに、特定のインターネットサイトへのトラフィックがまずは低調に始まり、

広告、リンク等により、短時間に指數的に増大することがある。このような需要増大のタイミングは予測が難しいことが多く、いかなるときにも需要を満たすために必要な計算および/またはネットワーク資源容量の予想または計画を難しくしている。

【0003】 1つの解決策は、通常の状態に必要であると推測されるよりも多くの資源容量（たとえば、より多くのCPU、より多くのサーバ等）を設けることである。しかし、このような資源の過剰割り当ては、費用が高くなり、必要に応じて容量に対する現行の方法では、追加資源のアクティブ化に手動での介入が必要である。すなわち、アクティブ資源が容量に達する前に、ユーザがバンダーとコンタクトをとり、追加資源を「ア

2

ンロック」してアクティブ化するためのコードを得なければならない。したがって、管理者は、アクティブ資源を監視し、必要に応じて予備資源をアクティブ化するために、動作中いかなるときにも（すなわち、1日24時間毎日）待機していなければならない。さらに、追加資源を即座に配備する必要がある需要が突然の上昇は、追加資源のアクティブ化が可能になる前に、サービスの中断を引き起こす可能性がある。

【0004】

【発明が解決しようとする課題】 したがって、本発明の主な目的は、動的にすなわち自動的に、ユーザ需要の変動に応じるために必要な資源容量を提供することである。本発明の別の目的は、需要を満たすために、追加の資源容量が必要な場合に、予備資源を自動的にアクティブ化することである。

【0005】

【課題を解決するための手段】 上記目的を達成するにあたり、本発明者らは、いくつかのアクティブ資源の負荷を監視し、いくつかのアクティブ資源の負荷が資源使用方針において指定された閾値に見合う場合に、予備資源を自動的にアクティブ化する装置および方法を考案した。

【0006】 本発明は、コンピュータ読み取り可能記憶媒体（たとえば、ディスク、RAM、または他のメモリ）に格納されるソフトウェアまたはファームウェア（すなわち、プログラムコード）において具現することができる。ソフトウェアは、たとえば、アクティブ資源（またはアクティブ資源プール）および予備資源（または予備資源プール）に直接、ネットワークを介して、周辺装置を介してなどでリンクされたコンピュータプロセッサによって実行することができる。資源使用方針は記憶媒体またはメモリに格納され、少なくとも1つのタスクをトリガするための少なくとも1つの閾値を含む。たとえば、「アクティブ資源No. 1の使用率が10秒よりも長い間90%を超える場合、予備資源No. 1をアクティブ化する」として、閾値90%と、資源をアクティブ化するタスクとを資源使用方針において指定することができる。タスクは、ページャ警報の送信、アラームの発生等も含みうる。コンピュータ読み取り可能プログラムコードもメモリに格納され、いくつかの機能を行うために、プロセッサによってアクセスされる。これらの機能には、いくつかのアクティブ資源、たとえばCPUまたはサーバに対する負荷の監視、および監視された負荷（たとえば、利用パーセントおよび/または使用持続期間の回数として）を資源使用方針に指定されている閾値との比較が含まれる。資源使用方針が命令すると、対応するタスクが実行される。

【0007】 一例として、負荷が資源使用方針に指定されている閾値に見合う場合、プロセッサは、アラームを管理者のページャに送信して、アクティブ資源が閾値に

50

達したことを管理者に警告しても、自動的に予備資源をアクティブ化してもよい。好ましくは、異なるレベルの資源割り当てに、監視方針の階層（すなわち、指定された閾値および対応するタスク）を用いる。たとえば、
 「10秒間90%の使用率」という指定された閾値では対応するタスクは「ページャの警報」であり、「5秒間95%の使用率」という指定された閾値では対応するタスクは「追加CPUのアクティブ化」であり、「10秒間1つまたは複数のシステムコンポーネントの95%の使用率」という指定された閾値では対応するタスクは「システム全体のアクティブ化」である。あるいは、管理者および/またはプロセッサが、資源消費率、残っている予備資源の数等、任意の適した要因に基づいて、動的に閾値および/またはタスクを設定することが可能である。たとえば、監視された資源消費率が「1時間当たり1%のアクティブ資源消費が追加」である場合、プロセッサおよび/または管理者が次のタスクを「ページャの警報発生」に設定することができる一方、資源消費率が「1分当たり1%のアクティブ資源消費が追加」である場合、プロセッサが次のタスクを「予備資源を自動的にアクティブ化」に設定することができる。

【0008】予備資源がアクティブ化されると、プログラムコードが、予備資源がアクティブ化されたという指示を生成する。上記指示は、ページャまたは電子メールの警告の管理者への送信、信号の負荷平衡器または他の周辺装置への送信、構成プロファイルの更新等をトリガすることができる。好ましくは、プログラムコードは、プロセッサまたは負荷平衡器等の周辺装置に、アクティブ資源とアクティブ化された予備資源との間で負荷を平衡させもする。一実施形態において、プロセッサまたは周辺装置は、負荷の少なくとも一部をいくつかのアクティブ資源からアクティブ化された予備資源に転送もし、これによって、アクティブ資源に対する負荷を低減する。好ましくは、プロセッサまたは周辺装置はまた、各アクティブ資源、アクティブ化された各資源、および各予備資源の構成プロファイルを維持する。構成プロファイルを用いて、利用可能な資源を脱断割り当てることができると共に、システムが過大割り当て状態（すなわち、予備資源が需要の処理に不十分の場合）に達する前に、システム管理者に警告することができ。さらに、負荷が指定された閾値（すなわち、アクティブ化閾値または別個の閾値全体）に見合うと、プロセッサが予備資源を非アクティブ化することで動作コストを低減し、予備資源を予備資源プールに戻すことが好ましい。次に、再度さらなる割り当てに資源が利用可能になる。本発明の別の実施形態は、監視およびアクティブ化に料金を課すことを含む。

【0009】したがって、本発明の装置および方法は、不必要かつコストのかかる資源の過大割り当てを行うことなく、ユーザの需要の変動に見合うように、需要に

じた資源容量を提供する。さらに、予備資源は自動的にアクティブ化されるため、手動での介入が必要なく、そのためサービスの中断および管理者が動作中に待機する必要性を低減するか、なくす。しかし、本発明はまた管理者に警告して、所望であれば、手動での介入および/または監視する機会を管理者に与えることが好ましい。もはや不必要な場合にアクティブ化された予備資源を非アクティブ化することができるように、またはそうして追加された予備資源を必要な場合にアクティブ化することができるように、予備資源をアクティブ化した後も、負荷を監視し続ける。したがって、予備資源は、自動的に予備資源プールに戻され、それによってコストが低減され（すなわち、ユーザは資源が割り当てられている時間についてのみ支払う必要がある）、かつ/または、効率の最も適した資源を割り当てるために他のシステムが予備資源を利用できるようにする。さらに、本発明はまた、購入者が前もってサービスについて支払うが、実際に必要になるまでハードウェアについては支払わない、ビジネスを行うための新しいユーティリティモデルを含む。

【0010】本発明のこれらおよび他の重要な利点および目的は、添付の説明、図面、および特許請求の範囲においてさらに説明されるか、これから明白になるであろう。

【0011】本発明の例示的かつ目下好ましい実施形態を図面に示す。

【0012】

【発明の実施の形態】図1は、本発明の監視およびアクティブ化装置10が使用可能な各種ハードウェアコンポーネントを示す。装置10は、プロセッサまたはCPU（中央演算処理装置）20と、コンピュータ読み取り可能記憶装置またはメモリ30とを備える。オプションとして、プロセッサ20は、ページャ、電話、コンピュータ等に従来の通信ネットワークまたは他のネットワークリンクを介して直接リンクすることができるか、または、イベントマネージャ50を介して上記に間接的にリンクすることができる。プロセッサ20はまた、負荷平衡器等のオプションの周辺装置40を通してアクティブ資源60（たとえば、CPUまたはサーバ）または複数のアクティブ資源60、61、62を内部に有するアクティブ資源プール70にもリンクされる。プロセッサ20はまた、同じまたは別のオプションの周辺装置40を通して予備資源80または複数の予備資源80、81、82を内部に有する予備資源プール90にもリンクされる。コンピュータ読み取り可能プログラムコード（たとえば、メモリ30に格納され、プロセッサ20によりアクセスされる）は、アクティブ資源60（またはプール70）に対する負荷を監視し、監視された負荷をメモリ30に格納可能な資源使用方針310（図3）に指定された閾値315（図3）と比較するために設けられる。

5

監視された負荷が指定された閾値315に見合う場合、プログラムコードが、指定された閾値315に対応するタスク320(図3)のパフォーマンスを(たとえば、プロセス20、周辺装置40等により)命令する。たとえば、対応するタスク320は、イベントマネージャ50へ通知すること、または予備資源プール90からの予備資源80をアクティブ化することでありうる。予備資源80がアクティブ化されると、プログラムコードが、予備資源80がアクティブ化されることの指示の生成を命令することが好ましい。指示は、インターネット10、HTML(ハイパーテキストマークアップ言語)ページ、ページ、電子メール等を介して管理者に送信される警告であることができる。指示はまた、周辺装置40に送信される信号、または警告と信号の組み合わせでもあってもよい。好ましくは、プログラムコードはまた、プロセス20(または周辺装置40)にアクティブ資源60(またはプール70)とアクティブ化された予備資源80との間で負荷を平衡するよう命令する。負荷が指定された閾値315未満に下がる(または、別個に指定された閾値315に見合う)と、プログラムコードはまた、予備資源80を非アクティブ化させ、予備資源プール90に戻すように命令することが好ましい。また、好ましい実施形態では、構成プロファイル(たとえば、各資源およびその状態を含む、メモリ30内のデータベース)が、アクティブ資源60および予備資源80(およびオプションとしてプール70、90)の状態を追跡するために更新(またはさらに維持)される。

【0013】プロセス20は、Intel PENTIUM(登録商標)等任意の従来のプロセッサ、デスクトップまたはラップトップコンピュータ等のパーソナルコンピュータシステム(PC)であるか、または本発明と併用するために特別に製造されたASIC(特定用途向け集積回路)であることが理解されよう。同様に、コンピュータ読み取り可能記憶媒体またはメモリ30は、RAM、ROM、磁気テープ、コンパクトディスク、これらの組み合わせ等、任意の適したメモリであることができる。プロセス20およびメモリ30は、別体のユニットである必要はなく、組み合わせでもよく、あるいは別個に収納して、互いに遠隔ネットワークまたは他の適した接続を介してリンクしてもよい。同様に、インターネット、イントラネット、LAN、WAN等を介して接続またはリンクされるいくつかの(すなわち、1つまたは複数の)プロセス20および/またはいくつかの(すなわち1つまたは複数の)メモリ30があってもよい。このような場合、上記プログラムコードの格納を、前記メモリにわたって分散することができ、または各種プロセスによって実行することができる。

【0014】イベントマネージャ50は、システム資源を監視し、監視要求を構成し、資源の状態をチェックし、構成された条件に見合う場合には通知を送信する。

6

イベントマネージャ50の一例は、「Event Monitoring Service User's Guide」(HP Part Number B7612-9001, 5, November 1999, Copyright 1999 Hewlett-Packard Company)に記載されている。また、イベントマネージャ50は、イベント通知を任意の適した装置に渡して、指定された閾値315(図3)に見合ったことを管理者に通知することができることも分かる。さらに、イベントマネージャ50は、イベント通知を渡して、指定された閾値315(たとえば、CPUがアクティブ化された)に見合ったことをベンダーに通知することで、ベンダーは顧客が課金方針に合致しているかを監視し、予備資源80がアクティブ化される場合にユーザに料金を課すことなどをを行うことができる。たとえば、イベントマネージャ50は、イベント通知を従来のページ、セルラ電話または他のハンドヘルド(携帯)電話、インターネットアプライアンス、デスクトップまたはラップトップコンピュータ、従来のディスプレイ、信号光、オーディオスピーカ、インターネットHTMLページ、電子メールアドレス、それらの組み合わせ等を通じて行うことができる。同様に、イベントマネージャ50は、プロセス20に接続して接続することでもでき、遠隔接続を介して(たとえば、インターネット、通信ネットワーク、モデム接続、T-1、またはDSL接続、衛星等を介して)プロセス20がアクセスしてもよい。これらの例は単なる例示であり、本発明の範囲を制限する意図がないことが明白に分かるであろう。

【0015】図2は、図3のシステムにおいてアクティブであるか、または予備でありうる各種タイプの資源60、80を示す。アクティブ資源60または予備資源80は、CPU、メモリ、周辺装置等の個々のコンポーネント200〜208および211〜219であり、かつ/またはホストコンピュータまたはPC、サーバ等のシステム210、225全体であることができる。同様に、アクティブ資源60および予備資源80は、CPU、メモリ、およびPCシステム等、それらの組み合わせであってもよい。プログラムコードが命令するように、いくつかの(すなわち1つまたは複数の)資源60、80を監視し、資源60、80に直接接続されるか、または従来のネットワーク220を介して資源60、80に接続されたプロセス20で実行することができる。

【0016】図1および図2に示す構成は単に例示であり、本発明の教示を制限する意図がないことが分かる。たとえば、資源60、80、およびプロセス20(および関連するメモリ30、周辺装置40、およびイベントマネージャ50)をクラスタ化する必要はない。すなわち、資源プール70、90、またはネットワーク220にわたって分散された資源60、80を含むことができる。同様に、プロセス20および/または周辺装置40は、1つ以上の資源プール70、90における

資源60、80を制御（たとえば、監視、アクティブ化、非アクティブ化）することができる。

【0017】図3は、（たとえば、メモリ30に格納された）監視方針を示す。監視方針300は、閾値315を指定すると共に、それに対応するタスク320を有する少なくとも1つの資源使用方針310を含む。しかし、監視方針は、それぞれ1つまたは複数の閾値315〜317を指定する任意の数の資源使用方針310〜312（方針1、方針2、...、方針n）と、これに割り当てられるか、または対応する任意の数の対応するタスク320〜323（タスク1A、タスク1B、タスク2、...、タスクn）を含むことができる。さらに、1つ以上の資源使用方針310が同じタスク320（図示せず）をトリガすることができ、単一の資源使用方針310が1つ以上のタスク320をトリガすることができる（たとえば、タスク1Aおよび1Bの双方が図3における閾値1に対応する）。

【0018】好ましくは、監視方針は、閾値315およびタスク320を有する資源使用方針310の階層を含む。すなわち、アクティブ資源60に対する需要が高くなる初期の徴候を認識するために、低閾値を設定することができ、プロセッサ20は、これらの需要が高くなる初期の徴候を管理者に警告することができる。したがって、管理者は、初期段階で（すなわち、追加資源がアクティブ化され、それらの使用に料金が課される前に）手動で介入することが可能である。たとえば、管理者は、高い需要がユーザトラフィックの増大によるものか、または資源の過剰利用を引き起こしたソフトウェアールーチン（たとえば、永くループ）における障害によるものかをチェックすることができる。障害がある例では、管理者が30

＊が手動で介入して、アプリケーションに起因する過剰利用をリセットすることが可能である。または、管理者が、次に指定されている閾値に対応するタスク320によって指定される高いコストの資源の代わりに低コストの資源を割り当てるよう決定する別の例では、管理者が、手動でこれらの代替資源を割り当てることができる。より高い閾値も設定可能であるため、警告を管理者に送信した後に資源の利用が増大し続ける場合に、プログラムコードが予備資源80のアクティブ化を命令し、所望であれば、メッセージを管理者に送信して、予備資源80がアクティブ化されたことを通知することができる。このように、管理者は、予備資源80がエラーでアクティブ化されたのか、またはテストの一部としてアクティブ化されたのかをチェックし、必要であれば（たとえば、料金を発生させる前に）予備資源を非アクティブ化することが可能である。負荷が増大し続ける際にシステム全体をアクティブ化するために、別のさらに高い閾値を指定することが可能である。追加資源がない（すなわち、システムが過剰割り当て状態に入っている）場合、さらに高い閾値を指定して、アクティブ資源60およびアクティブ化された予備資源80の組み合わせとすることが可能である。このさらに高い閾値が見合う場合、プログラムコードは、優先度の高い警告を管理者に送信するよう命令して、負荷が増大し続けており、アプリケーションを閉じるなどの手動の介入がなく、資源が不足しうることを管理者に通知することができる。このような階層の一例を以下の表に示す。

【0019】

【表1】

閾値	タスク
1) 平均CPU使用率>10 秒間 90%	1) 「高CPU使用率」アラームの送信
2) 平均CPU使用率> 5 秒間 95%	2) 予備資源のアクティブ化および「高CPU使用率-予備アクティブ化」のアラーム送信および負荷平衡化
3) 平均組み合わせCPU使用率> 10 秒間 80%	3) 予備資源のアクティブ化（全システム）
4) 平均組み合わせCPU使用率> 10 秒間 80%および追加予備資源利用不可	4) 「高CPU使用率-さらにアクティブ化する資源なし-手動介入必要」アラームの送信

【0020】閾値315は、1) 監視前に設定する（すなわち、資源使用方針310で指定する）、2) 監視中に設定または変更する、または3) その組み合わせに基づいて設定してもよいことが明明白に分かる。すなわち、閾値は、負荷が閾値に達する前に、資源使用方針310で指定することができる。しかし、プログラムコードは、監視中に閾値315を命令することができるか、

または管理者が手動で監視中に閾値315を指定することができるか、またはそれら双方が可能である。すなわち、閾値315には、資源消費率、残っている予備資源の数、または任意の他の適したアルゴリズム等、各種要因に基づいて対応するタスクを指定しかつ／または割り当てることができる。

【0021】一例として、資源消費率が「1時間当たり

1%のアクティブ資源消費が追加」の場合、タスク「ページ警報を発生」を次に指定された閾値315に割り当てることができる一方、資源消費率が「30秒当たり1%のアクティブ資源消費が追加」である場合、タスク「予備資源の自動アクティブ化」を次に指定された閾値315に割り当てることができる。別の例では、指定された第1の閾値315に達すると、プログラムコードが、その状態を管理者に送信するように命令し、次に管理者が、その状態および/または他の考慮事項に基づいて、第2の閾値315および対応するタスク320を指定することができる。さらに別の例では、管理者が、監視前、または監視中に、キーボード入力、電子メール、FTP（ファイル転送プロトコル）、音声コマンド等を経て、1つまたは複数の資源使用方針310を送信することができる。実際に、管理者は、監視前または監視中のいつでも監視方針全体を送信することができる。同様に、資源使用方針310は、閾値315およびタスク320を含め、デフォルトとして監視方針300において確立可能である。

【0022】図4は、アクティブ資源60に対する負荷を監視し、負荷が資源使用方針310で指定された閾値315に見合う場合に、対応するタスク320を実行する方法を示す。ステップ400において、プログラムコードが、アクティブ資源60に対する負荷の監視を命令する。たとえば、アクティブ資源60がCPUまたはサーバである場合、負荷は経時に変化する使用率であることができる。または、たとえば、アクティブ資源60がメモリである場合、負荷は残っている空間に対する空間利用率であることができる。負荷の監視に適した任意のアルゴリズムを、本発明の教示の下で使用することが可能であり、ソフトウェアを監査する従来のコンポーネントまたはシステム（たとえば、Microsoft Windows（登録商標）Resource Meter）を用いて行うことができる。プログラムコードはまた、ステップ410において、監視された負荷を資源使用方針310に指定された閾値315と比較するように命令する。たとえば、閾値315（たとえば、90%、120MB、95%等）を「 $\geq 90\%$ 連続した10秒間のCPU利用率」「 $> 120\text{MB}$ RAM利用率」「 $< 95\%$ ポート可用性」または任意の他の適した指定閾値として指定することができる。ステップ420において、監視された負荷が指定された閾値315に見合わない（たとえば、過去10秒間にCPUの利用率が90%未満である）場合、ルーチンがステップ400に戻り（422）、プログラムコードがアクティブ資源60に対する負荷の監視を継続するように命令する。しかし、ステップ420において、監視された負荷が指定された閾値315に見合う（たとえば、過去10秒間にCPUの利用率が90%以上である）場合、ルーチンは進み（424）、資源使用方針310が対応するタスク320を行うように命令する。図

4は、424の後に続く対応するタスクの例示的な3つの選択肢を示しており、これらについてより詳細に以下に説明する。

【0023】一例として、タスク320は、監視されたアクティブ資源60に対する負荷が「 $\geq 90\%$ CPU利用率」である場合に、「アラーム生成」として定義することができる。したがって、監視された負荷が、資源使用方針310に指定された閾値315に見合う場合、プログラムコードは、ステップ430において、イベントマネージャ50に送信されるアラームを生成するように命令する。たとえば、プロセス320は、管理者にアクティブ資源60の状態を警告するページを送信する。ルーチンはステップ400に戻り（435）、プログラムコードがアクティブ資源60の監視を継続するように命令する。

【0024】別の例として、タスク320は、監視されたアクティブ資源60に対する負荷が「5秒間少なくとも90%のCPU利用率」である場合に、「予備CPUのアクティブ化」として定義することができる。したがって、監視された負荷が指定された閾値315に見合う場合、プログラムコードは、ステップ440において、制御ソフトウェアおよび/またはハードウェアを用いて、予備資源80をアクティブ化するように命令する。好ましい実施形態では、各予備資源80がブート時にオペレーティングシステム（OS）によって電気的にアクティブ化され、認識される。このようにして、オペレーティングシステムは、予備資源80に関する状態データを得ることができ、システムをリポート等することなく、予備資源80をテストし、アクティブ化することなどができる。しかし、予備資源80は、アクティブ化前には利用されない。予備資源80を機能的にアクティブ化するために、OSは、資源を利用可能なまたはアクティブ化された予備資源として認識するよう命令される。ユーザが所定のパスコードを入力すると、ハードウェアを機能的にアクティブ化するアクティブ化API（アプリケーションプログラマインタフェース）ソフトウェアが、現在利用可能となる。すなわち、管理者が、追加容量をシステムに追加しなければならないと認識すると、ベンダーとコンタクトをとり、パスコードを受け取る。次に、管理者は、バスコードをアクティブ化APIに入力し、次に資源が「アンロック」される。したがって、予備資源80をオンラインにしなければならぬことを検出する（すなわち、ステップ400における監視中）と、予備資源を自動的にアクティブ化する本発明の実施形態では、プログラムコードは、メモリ30に格納されたバスコードを「ルックアップ」するか（たとえば、ルックアップテーブルまたは他のデータ構造において）、あるいはベンダーとの接続を確立させ（たとえば、モデムダイヤルアップ接続、DSL等）、バスコードをダウンロードする、そうでなければ検索すること

11

ができる。次に、プログラムコードは、パスコードをアクティブ化APIに提示し、予備資源80をアクティブ化する。適切な閾値315は、パスコードを獲得して提示する際の遅延を考慮に入れて指定することができる。しかし、別の実施形態では、パスコードは必要なく、プログラムコードが行う他のアクティブ化要求または信号に応答するように、アクティブ化APIをプログラムしてもよい。アクティブ化APIをOS、OSへのオペレータなどに統合してもよいことが分かる。あるいは、予備資源80は、スタートアップ時に電氣的に非アクティブ化して、使用前に電氣的にアクティブ化させてもよい。電氣的に非アクティブな資源をアクティブ化するためのプログラムコードまたはファームウェア（たとえば、コンピュータBIOS、またはOSパワー管理能力）もまた、当業者の理解の範囲内である。

【0025】ステップ442において、プログラムコードは、予備資源80がアクティブ化されたことを指示するよう命令する。指示は、プロセッサ20から管理者にイベントマネージャ50を介して送信されるメッセージ、周辺装置40に送信される信号、プログラムコード内のサブルーチンに送信される信号、これらの組み合わせ等でありうる。ステップ447において、プログラムコードが、アクティブ資源60とアクティブ化された予備資源80の間で負荷を平衡するように命令する。たとえば、負荷平衡器40が、新しいタスクをアクティブ化された予備資源80にルーティングすることができる。あるいは、負荷平衡器40が、アクティブ資源60からの負荷の少なくとも一部をアクティブ化された予備資源80に転送してから、処理要件、時間要件等に基づいて、新しいタスクをより良好に処理することのできる資源60または80に新しいタスクをルーティングする。ステップ460において、プログラムコードは、構成プロファイル500（図5）を更新するように命令する。構成プロファイル500は、図5に示すもの等、（たとえば、510において）指定された各資源60、80の状態520（たとえば、アクティブ、非アクティブ、故障、容量等）を含む、メモリ30に格納されるデータベースであることができる。構成プロファイル500は、アクティブ資源60および予備資源80（アクティブ化されたものと非アクティブの双方）を少なくとも追跡するために使用可能である。このように、負荷平衡器40は、利用可能な資源60、80の間で負荷をルーティングまたは平衡して、必要な場合には追加予備資源80をアクティブ化し、それ以上利用可能な予備資源80がないなどの場合にアラームを発生することができる。

【0026】さらに別の例として、タスク320は、監視されたアクティブ資源60および/またはアクティブ化された資源80に対する負荷が「15秒間CPUの利用率が85%未満」（すなわち、指定された閾値31

12

5）である場合に、「予備CPUの非アクティブ化」として定義することができる。したがって、監視された負荷が閾値315に見合う場合、プログラムコードは、ステップ450において（図4）、オプション的に、アクティブ化された予備資源80に対して残っているすべての負荷をアクティブ資源60に戻し、ステップ455において、再び、予備資源のアクティブ化に関して上述した制御ソフトウェアおよびハードウェアを用いて、予備資源80を非アクティブ化するように命令する。このように、予備資源80は予備資源プール90に戻され、コスト節約のためにパワーダウンしたり、他のシステムに利用可能になったりなどすることができる。

【0027】非アクティブ化タスク320に対応する指定閾値315は、アクティブ化タスク320に対応する指定閾値と同じ指定閾値315であってもよいことを理解されたい。しかし、好ましくは、アクティブ資源60に対する負荷が指定された閾値315付近で変動するときに、パウンス（すなわち、予備資源80のオンおよびオフ）を低減するために、非アクティブ化タスク320に対応する指定閾値315は別個の指定閾値315である。たとえば、予備資源80を非アクティブ化する前に、相当低下しなければならないように、非アクティブ化タスク320に対応する指定閾値310が設定される。プログラムコードは、ここでは、ステップ460において、予備資源80が非アクティブ化され、予備資源プール90に戻ったことを示すように、構成プロファイル500を更新するよう命令する。

【0028】図4に示すルーチンにおける424において、プログラムコードが任意の適したタスク320、閾値315、または資源使用方針310を命令可能なことを理解されたい。たとえば、プログラムコードは、アクティブ資源60が不足し始める（たとえば、閾値「処理エラーが戻された」が見合う）場合、またはアクティブ資源60をアップグレードすべき（たとえば、閾値「処理速度がXMHz未満」に見合う）場合、予備資源80をアクティブ資源60と交換するように命令することができる。さらに、好ましい実施形態によれば、プログラムコードは、タスク320を行った後に、さらなる負荷の変更を見ずとも検出し（すなわち、指定された閾値315を見ずとも）、それに従って作用する（すなわち、対応するタスク320が行われる）ように、負荷の監視を継続するよう命令する。

【0029】同様に、コンピュータ読み取り可能プログラムコードは、任意の適したコンピュータ読み取り可能なプログラミング言語を用いて、従来通りにプログラム可能であり、1つまたは複数の関数、ルーチン、サブ関数、およびサブルーチンを含んでもよく、単一のソフトウェアパッケージに組み合わさる必要がないことを理解されたい。

【0030】本明細書で用いる「見合う」および「見合

13

った」という語（たとえば、ステップ420においては、監視された値または対応の閾値に達するか、閾値を超えるものとして定義されることもまた、理解されたい。しかし、閾値315は、負荷が「等しい」、「超過する」、「未満」、「以上」、「略称にわたり平均化する」、「時間間隔当たりで生じる」等でなければならない条件を含むように、資源使用方針310によって特定することができる。

【0031】図4に示すステップを図示の順序で実行する必要がないことも、さらに理解されたい。たとえば、ステップ430においてアラームを生成した後に、ステップ440において予備資源をアクティブ化（440）してもよい。同様に、本発明はまた、図4に示すステップよりも少数のステップを含む方法、および図4に示すステップよりも追加のステップを含む方法も意図する。たとえば、424にさらなるタスク320を含めてもよく、ステップ455において非アクティブ化する予備資源80に対する負荷がない場合、ステップ450を省いてもよい。

【0032】本発明の別の実施形態は、自動監視およびアクティブ化価格付けを含む。一実施形態において、基本料金がシステム（たとえば、PC、サーバ等）に課される。基本料金には、通常ベースで使用されると予期される資源（すなわち、アクティブ資源60）の料金が含まれる。そして、ユーザには、3つの異なる課金オプションのうち少なくとも1つを用いて、予備資源80の使用に料金を課す。1つのオプションは、基本料金に加え、予備資源80がアクティブ化されると課される「一度限り」の使用料金を課すことを含む。このオプション下では、予備資源80が一旦アクティブ化されると、予備資源80が購入され、後に使用される分には追加料金が課されない。第2のオプションは、基本料金に加え、「必要に応じた」料金を課すことを含む。このオプション下では、「必要に応じた」料金が、アクティブ化された予備資源80の使用期間中のみ課される。この場合、予備資源の基本料金は本質的にオプション料金である。予備資源80が一旦非アクティブ化され、ユーザプール90に戻されると、追加料金は無い。第3のオプションは、予備資源80をアクティブ化したときのみ料金を課すことを含む。このオプション下では、基本料金は課されず、第2のオプションに関して上述した「必要に応じて」ベースで課されるだけである。これらの課金オプションは、監視方針に含め、プログラムコード（たとえば、タスク320として）で命令することができる。たとえば、方針で「負荷>90%である場合、資源をアクティブ化し、課金エージェントに資源を使用中であることを警告する」と述べてもよい。上記3つの例は単なる例示であり、本発明の下で、管理者への警告に料金を課すなど（しかし、これに限定されない）他の実施形態も意図される。

14

【0033】本発明の例示的かつ目下好ましい実施形態を本明細書に詳細に説明したが、本発明の概念は、別様に様々な具現かつ採用することができ、特許請求の範囲は、従来技術によって制限されるものを除き、かかる変形を含むように解釈されるものであることを理解されたい。

【0034】本発明の態様を以下に例示する。

【0035】1. 予備資源（80）を自動的にアクティブ化する方法であって、いくつかのアクティブ資源（60）に対する負荷を監視するステップと、前記負荷を資源使用方針（310）に指定されている閾値（315）と比較するステップと、前記資源使用方針（310）によって命令される場合に、前記予備資源（80）を自動的にアクティブ化するステップと、を含む、方法。

【0036】2. 前記予備資源（80）をアクティブ化した後に、前記資源使用方針（310）を更新するステップをさらに含む、上記1記載の方法。

【0037】3. 前記負荷を前記いくつかのアクティブ資源（60）と前記アクティブ化された予備資源（80）の間で平衡するステップをさらに含む、上記1記載の方法。

【0038】4. a) 前記いくつかのアクティブ資源（60）および前記アクティブ化された予備資源（80）に対する組み合わせられた負荷を監視するステップと、b) 前記組み合わせられた負荷を第2の資源使用方針（310）に指定されている第2の閾値（315）と比較するステップと、c) 前記第2の資源使用方針（310）によって命令される場合に、前記いくつかのアクティブ資源（60）および前記アクティブ化された予備資源（80）から選択された資源を非アクティブ化するステップと、d) 前記選択された資源が非アクティブ化されたことを示すステップと、をさらに含む、上記1記載の方法。

【0039】5. 前記監視された負荷に基づいて、前記資源使用方針（310）によって命令されるように、イベントマネージャ（50）に通知するステップをさらに含む、上記1記載の方法。

【0040】6. 前記資源使用方針（310）は、前記閾値（315）に見合う場合には、いくつかのアラーム（320）をトリガし、前記資源使用方針（310）は、前記いくつかのアラーム（320）への応答が受信されない場合、前記予備資源（80）をアクティブ化するように命令する、上記1記載の方法。

【0041】7. 予備資源（80）を自動的にアクティブ化する装置（10）であって、少なくとも1つのコンピュータ読み取り可能記憶媒体（30）と、前記少なくとも1つのコンピュータ読み取り可能記憶媒体（30）に格納される資源使用方針（310）と、前記少なくとも1つのコンピュータ読み取り可能記憶媒体（30）に格納されるコンピュータ読み取り可能プログラムコード

と、を備え、該コンピュータ読み取り可能プログラムコードは、a) いくつかのアクティブ資源(60)に対する負荷を監視するためのプログラムコード、b) 前記監視される負荷を資源使用方針(310)に指定されている閾値(315)と比較するためのプログラムコード、およびc) 前記資源使用方針(310)によって命令される場合に、前記予備資源(80)をアクティブ化するためのプログラムコード、を含む、装置。

【0042】8. 階層資源配置方式に基づいて、アクティブ化する予備資源(80)を選択するためのプログラムコードをさらに備える、上記7記載の装置。

【0043】9. 予備資源プール(90)から前記予備資源(80)を選択するためのプログラムコードをさらに備える、上記7記載の装置。

【0044】10. 前記少なくとも1つのコンピュータ読み出し可能記憶媒体(30)に格納される構成プロファイル(500)と、前記資源(60、80)それぞれの利用可能性を追跡するように、前記構成プロファイル(500)を更新するためのプログラムコードと、をさらに備える、上記7記載の装置。

【0045】

【発明の効果】本発明によれば、動的にすなわち自動的に、ユーザ需要の変動に応じるために必要な資源容量を提供することができる。さらに、需要を満たすために、

追加の資源容量が必要な場合に、予備資源を自動的にアクティブ化することができる。

【図面の簡単な説明】

【図1】監視およびアクティブ化装置のコンポーネントのブロック図である。

【図2】監視およびアクティブ化装置と併せて用いるための資源のプールを示す。

【図3】本発明の装置および方法と併せて用いるための監視方針を示す。

【図4】アクティブ資源に対する負荷を監視するステップと、負荷が閾値に見合う場合に対応するタスクを実行するステップとを示すフローチャートである。

【図5】例示的な構成プロファイルである。

【符号の説明】

10 監視およびアクティブ化する装置

30 コンピュータ読み取り可能記憶媒体またはメモリ

50 イベントマネージャ

60 アクティブ資源

80 予備資源

90 予備資源プール

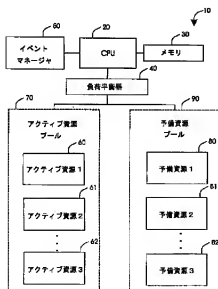
310 資源使用方針

315 閾値

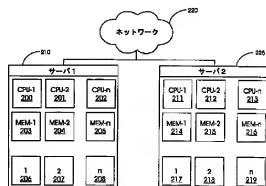
320 タスク

500 構成プロファイル

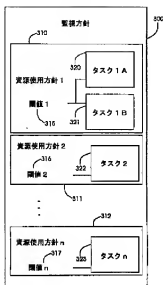
【図1】



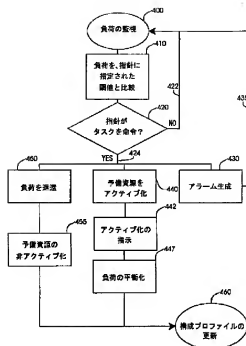
【図2】



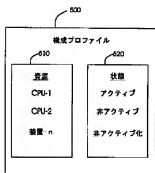
【図3】



【図4】



【図5】



フロントページの続き

(72)発明者 トーマス・イー・トゥリッチ・ジュニア
アメリカ合衆国75206テキサス州ダラス、
マッコマス・プールバード 5903

Fターム(参考) 5B045 G304
5B089 GA11 GA21 GB02 JB16 JB18
KA06 MA07
5B098 AA10 GA04 GD02 GD14